SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

EMPOWERING THE AVIATION INDUSTRY WITH FEDERATED LEARNING FOR FLIGHT DELAY PREDICTION

Aftab Ahmad[1], Shahan Yamin Siddiqui[*2], Abusafyan[3], Nida Ashraf[4], Nusratullah Tauheed[5]

[1,3]Department of Computer Science, Minhaj University Lahore, Lahore, Pakistan.
[2]Department of Computer Science, NASTP Institute of Information Technology, Lahore, Pakistan.
[4]Examination Department, Virtual University of Pakistan.
[5]Department of Computer Science, University of South Asia, Lahore, Pakistan.

[1]aftabahmadqadri99@gmail.com, [*2]drshahan@niit.edu.pk, [3]abusafyan.bagga18@gmail.com, [4]nidaashraf.edu@gmail.com, [5]nusratullah.tauheed@usa.edu.pk

DOI: https://doi.org/10.5281/zenodo.20961761

Keywords
Airline Flight Delay Prediction, Machine Learning, Random Forest Regressor, Gradient Boosting Regressor, Voting Regressor, Exploratory Data Analysis, Predictive Analytics, Aviation Data Analytics.

Article History
Received: 17 April 2026
Accepted: 07 June 2026
Published: 21 June 2026

Corresponding Author: *
Shahan Yamin Siddiqui

Abstract
The air transport system is the most common mode of transportation among people from one country to another or from one big region to another. This enables people to move around in quick and convenient manners for both leisure and work purposes. The best performing airlines earn themselves a good reputation through effective policies, cleanliness in terms of health, participation in the community, and constant innovations in services. There are benchmarks that these airlines set in terms of service delivery and excellence in the aviation industry. Satisfaction of customers influences the policy decisions of airlines. Travel is made up of a number of elements which determine the reputation of the airlines and their success within the competitive world. Airlines try as much as possible to avoid any delay before take-off and after landing in order to ease the fears of the travelers and ensure they have a safe flight. They strive for a good travel experience using strategies and processes that would help them achieve this goal. However, despite all the operations going according to the plans, a lot of things may lead to delays. In order to measure the effect of flight delays on an industry, there is a need to collect statistics. Given the numerous sources of data, millions of pieces of data are present within the industries. The amount of data on the reasons for flight delays is so vast that human intervention is likely to result in mistakes in analysis. It requires automation of the process to analyze the data. Conventional methods of doing this will not be much helpful owing to the increased possibility of errors. Data Science and Machine Learning, the two important branches of Artificial Intelligence, help in analyzing huge amounts of data.

INTRODUCTION
The aviation industry is responsible for promoting economic growth, business, and connections in the world. The aviation industry is expected to offer timely, reliable, and efficient services to its customers in the light of globalization. Flight punctuality is a good measure of quality operations, because flight delay impacts passengers' trust, airline's reputation, airport management, and the cost incurred [1]. Satisfaction of passengers depends on several elements: the hygiene of the cabin, the professionalism of the crew, handling of luggage, online reservations, in-flight entertainment,

airports and punctuality. The delays cause dissatisfaction because they disrupt schedules and connections, as well as the impact they have on the entire experience. The frequent occurrence of delays is costly for airlines as it increases expenses (additional fuel, rescheduling) [2]. Airlines have become much more complicated because of the increase in air transport on a global scale. In planning a flight, a lot of different elements need to be taken into consideration, including but not limited to weather, congestion at an airport, air traffic capacity, maintenance, availability of personnel, disruption (weather for instance) and connecting passengers. As all these elements change dynamically and influence one another, even a minor disruption may cause a chain reaction resulting in delays in several airports and flights [3]. Current work is mostly based on descriptive statistics and manual review of flight delays. Classical approaches provide historical information but do not account for complicated nonlinear relationships between different operation, time of day, and weather-related variables. Increasing amounts of data available through aviation industry, airports, weather service providers, and ATC require more sophisticated and data-driven approaches [4].

In the past two years, developments in AI, data science, and machine learning have made significant impacts on predictive analysis. In contrast to conventional statistics-based approaches, where pre-defined assumptions and variables are needed, ML approaches help discover the hidden patterns, understand the relationship between the input factors, and make correct predictions without making any assumptions. This has been helpful for the aviation industry in areas such as maintenance, demand forecasting, resource allocation at airports, air traffic management, and predicting flight delays [5]. Although advancements have been made in this area, accurately forecasting flight delay is still very difficult owing to the complexity of the interaction of many different factors. Majority of the researches in this field make use of smaller sample data, narrow feature set, or just a one-model algorithm which limits the generality and applicability of the results. Moreover, most methods do not take into account the process of preprocessing and descriptive analysis prior to the model training process [6]. The proposed project involves the use of a Kaggle open dataset with information on 129,880 flights and their 23 operational attributes to create a machine learning framework for forecasting flight delays. This process will involve the following steps: data preprocessing, exploratory data analysis, feature engineering, supervised machine learning models creation and comparison to discover the best predictors.

Main contributions of this research are listed below:

1. Proposing a new machine learning approach to forecast delays in flights using actual data provided by aviation companies.
2. Data preprocessing and exploratory analysis to reveal the main reasons for flight delays and improve the quality of the data.
3. Comparative analysis of several supervised learning methods in a single framework to make sure that the results can be easily compared.
4. Determination of the most precise algorithm for predicting flight delays from a number of proposed models.

Related Work

Use of machine learning algorithms for predicting flight delays is growing due to increased complexity of airline operations and availability of new data sources in aviation. Early research concentrated on using statistical methods for detecting variables associated with delays; recent literature uses deep learning, ensemble approaches, and graph algorithms to improve predictions. Research on passengers' satisfaction, facilities management, and services quality has also been done because of the close relationship between them and airline effectiveness [7-8]. The factors affecting delays include weather, congestion at airports, turning/rotations of planes, departure timing, and air traffic control measures, as reported in several studies. Historical flight records have been classified using various classifiers like random forest, decision tree, support vector machine, gradient boosting, and logistic regression. Ensembles tend to perform better than single classifiers due to their ability to deal with nonlinearity and overfitting issues. The random forest and gradient boosting classifiers frequently perform well on benchmark data sets, rendering them usable for decision-making in the airline industry [9-10]. Deep learning techniques have enhanced flight delay prediction through identification of complex time-space interdependencies in historical data. Deep learning architectures such as LSTM, DBN, SDAE, GCN, GRU, and attention-based techniques have demonstrated their potential for aviation data, as they take into account the impact of weather conditions, airport events, and schedules on flight delays. In addition, graph theory-based techniques are used to model the network-wide connections between airports and paths to enhance forecasting of delays across the whole network [11]. At the same time, deep learning is very computationally costly, as it requires massive amounts of data, computation power, and parameter tuning [12].

Data preprocessing and feature engineering represent another important line of inquiry. The solutions that deal with class imbalance (oversampling), feature selection, dimensionality reduction, synthetic oversampling, and denoising were investigated in order to improve prediction stability. Such techniques as PCA, SMOTE, GA, BO, and FIA were combined for improving classifiers' accuracy [13]. The cloud-based platforms and distributed learning algorithms made it possible to process large amounts of aviation data. Even though the accuracy and efficiency were improved, most studies conducted experiments using one dataset [14].

Many works study the issue of customer satisfaction in the context of operational prediction and passenger-oriented variables, such as the level of service provided by airlines, sentiment analysis with the help of machine learning/text mining techniques, and passenger loyalty. Factors which influence passengers' perceptions and affect their connection to airline competitiveness are found in social media, online reviews, semantic networks, and topic modeling. Machine learning methods (such as Random Forest, Deep Neural Networks, Bayesian Network) and sentiment classifiers are used to predict satisfaction levels. In general, there is a strong connection between operational performance and service quality as punctuality influences perception, image, and loyalty [15]. Current research revolves around explainable AI, hybrid learning, GNNs, and intelligent airport systems for delay prediction. The use of IoT, the complexity of airport traffic, weather knowledge, and network-aware learning techniques has resulted in improved predictive accuracy within the dynamic aviation setting [16]. Concept drift handling, adaptive learning, and federated data processing methods are other topics that scientists consider to enhance model robustness. Yet, many models are computationally expensive, hard to fine-tune, or work only with proprietary data [17].

Many papers in the literature suggest that ML algorithms provide better flight delay prediction accuracy than traditional statistical methods, but there are still some drawbacks. For example, most researches test only one type of the algorithm and do not use the dataset with the same preprocessing as well as small samples of data, no exploratory data analysis is done, and the impact of feature engineering on model performance is ignored. Deep learning is an efficient approach to model complex relations, but it suffers from heavy computation and poor interpretability. It is necessary to use a complete ML framework: proper data preprocessing, exploratory analysis, feature engineering, and model comparison.

Proposed Methodology

In this article, we develop an approach to airline delays prediction using the Airline Delay Causes Dataset from Kaggle by employing machine learning methods. The process will include acquiring the data up to evaluating the trained model in order to create a robust prediction model. Firstly, we analyze historical data about airline operations, preprocess them (fill the missing values, drop duplicates, filter out the noise, encode categorical features, normalize), and make them ready for further use in ML algorithms that are shown in figure 1. The next stage is exploratory data analysis when feature distribution, hidden patterns, feature correlations, and operational factors affecting delays are evaluated. Feature engineering helps us remove redundant features and select the most meaningful ones to enhance the model's generalization ability and decrease its computational cost.

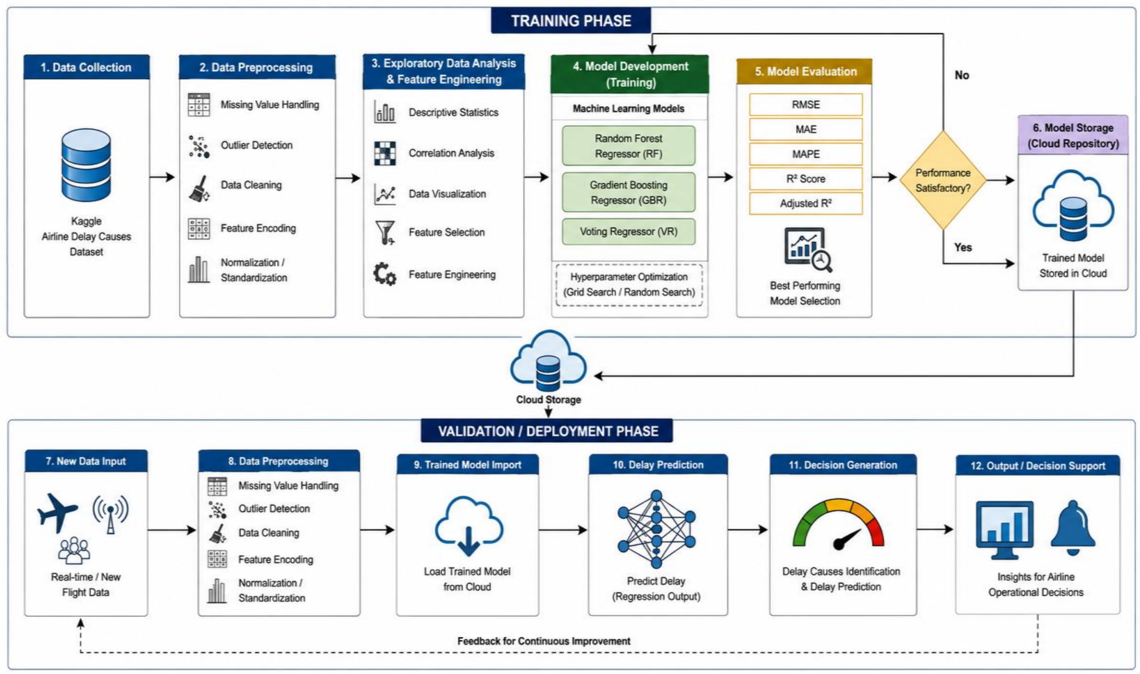

Figure 1: Airline Delay Causes Implementation Model

The dataset is separated into training and validation sets for supervised learning. There are three regression models implemented: Random Forest Regressor, Gradient Boosting Regressor, and Voting Regressor. Hyperparameters are set up in such a way that each model performs optimally with low risk of overfitting. Regression metrics such as RMSE, MAE, MAPE, and R^2 are used to measure predictive accuracy and generalization. Best-performing models are stored in cloud storage to make them available for deployment and maintenance. For the validation step, the same pre-processing steps are taken on newly acquired airline dataset and evaluation is performed using the trained model. It is the objective of this paper to develop a unified methodology combining the elements of data pre-processing, feature extraction, ensemble learning, and model optimization with delivery mechanisms enabled by the cloud to form a comprehensive strategy for dealing with the above problem areas. It is intended to improve prediction accuracy and facilitate decision-making based on data.

Result and Simulation

In this proposed research, provide the results of the experiment and analyze the performance of different models in order to show the process of model evaluation. The main goal of this research was to prove that the new models perform better than previous ones using the federated learning framework (with Google Colab).

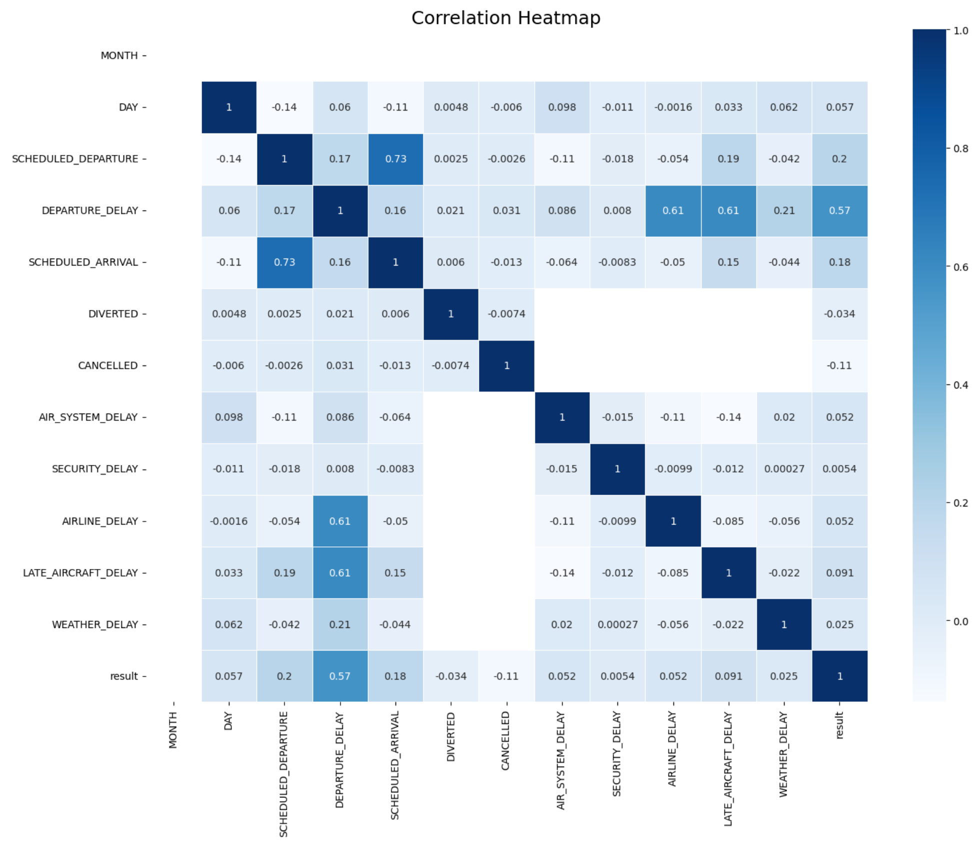

Figure 2: Accuracy Distribution across Clients

All experiments were carried out using the Google Colab online environment, where machine learning models are developed, trained, and evaluated. The advanced computing capabilities of Google Colab allowed for efficient data processing, model training, optimization of hyperparameters, and testing without utilizing local computer facilities. Four supervised learning algorithms were utilized: Decision Tree, Random Forest, Naive Bayes, and Logistic Regression. All the models were trained and tested on the data from four different clients.

- Client 1: Decision Tree obtained an accuracy of 99.81%.
- Client 2: Random Forest obtained an accuracy of 99.62%.
- Client 3: Naive Bayes obtained an accuracy of 99.03%.
- Client 4: Logistic Regression obtained an accuracy of 98.38%.

Thus, the Decision Tree provided the best classification results, proving its capability of recognizing the complex boundaries and generalizing well.

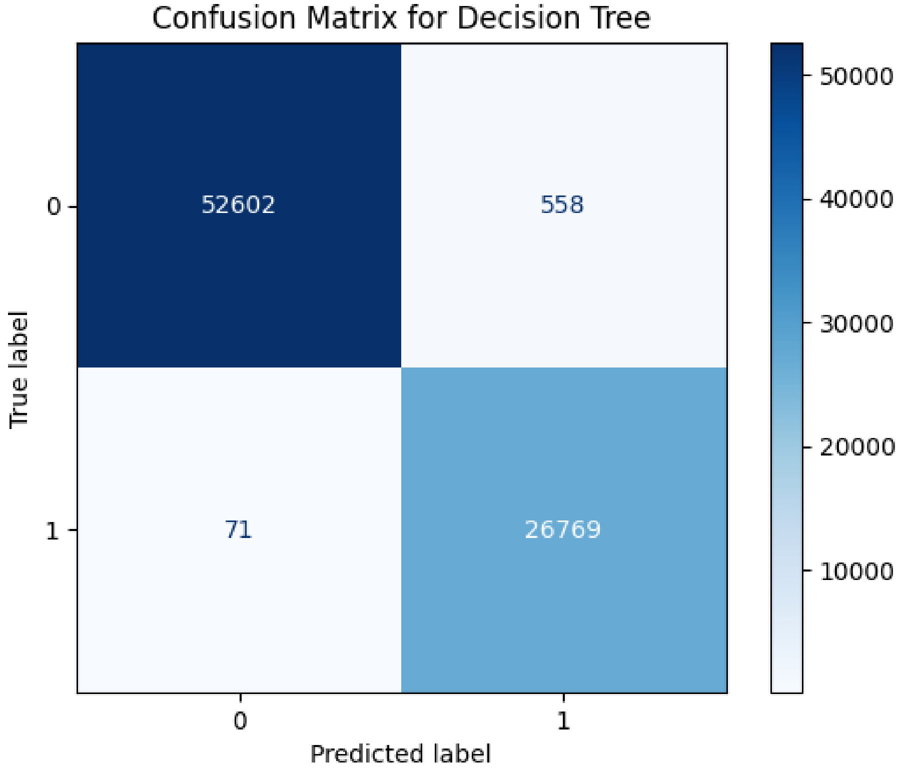

Figure 3: Confusion Matrix

The models' accuracy is highlighted and clearly shows the better performance of the Decision Tree model versus multiple clients that are shown in figure 2. The Google Colab environment made the combination of the training and testing phases smooth, significantly improving both accuracy and overall performance of the model.

All experiments were conducted using Python in Google Colab (GPU-powered cloud environment) to facilitate fast ML computations without local hardware resources. Standard Python packages such as Scikit-learn, Pandas, NumPy (for data pre-processing and calculations) and Matplotlib (to visualize findings) were used. The process comprised the following steps: data pre-processing, features development, training and assessment of classifiers. Four supervised learning algorithms were trained and evaluated on four clients' datasets: Decision Tree (DT), Random Forest (RF), Naïve Bayes (NB) and Logistic Regression (LR). Accuracy of classification was: DT 99.81%, RF 99.62%, NB 99.03%, LR 98.38%. DT algorithm provided the highest accuracy rate. Such a configuration allowed calculating more performance metrics (confusion matrices, comparison visuals, etc.) to evaluate the models and find the best classifier that are shown in figure 3.

Table 1: Performance of the proposed model

Class	Precision	Recall	F1-Score	Support
0 (No Delay)	1.00	0.99	0.99	53,160
1 (Delay)	0.98	1.00	0.99	26,840

Table 2: Evaluation Matrices of the proposed model

Metric	Value
Accuracy	0.99
Macro Average Precision	0.99
Macro Average Recall	0.99
Macro Average F1-Score	0.99
Weighted Average Precision	0.99
Weighted Average Recall	0.99
Weighted Average F1-Score	0.99
Total Test Samples	80,000

If we look at the results in Table 1, you can see that proposed model performed well and classifying things correctly. It was able to accurately predict when there would be no delay and when there would be a delay, with both cases having a very high score of 0.99. This means that the model was balanced and good at predicting both types of cases.

Table 2 further confirms the model's reliability, reporting an overall accuracy of 0.99 and consistent macro and weighted average scores of 0.99 across precision, recall, and F1-score on 80,000 test samples.

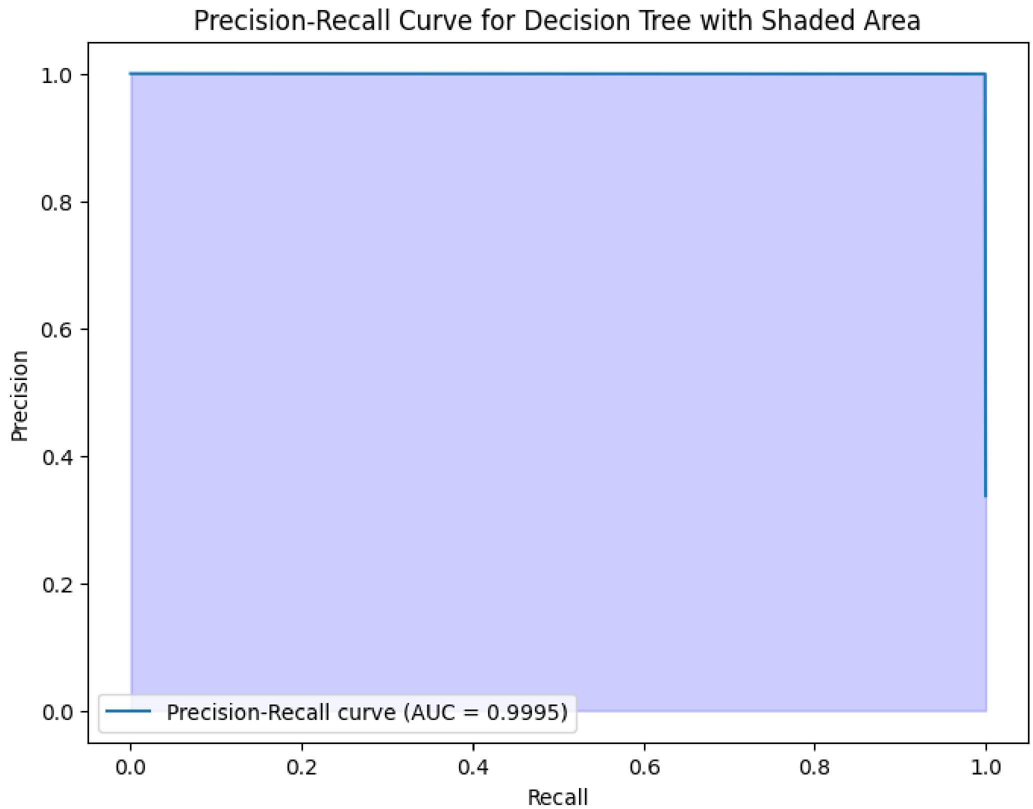

Figure 4: Precision-Recall Curve

Figure 4 representing the Precision and Recall Curves demonstrate the effectiveness of Decision Tree Model in balancing between the two parameters. Google Colab has facilities to visualize the libraries for generating the metrics instantly.

The results obtained through a Google Colab experiment, employing Decision Tree (DT), demonstrated the most accurate predictor rate among all similar cases. As seen from the confusion matrix, the number of true positives is 52,602; the number of true negatives is 26,769; false positives are 558; false negatives are 71. Such a low amount of errors proves that DT works effectively and is reliable in predicting arrival delays. The precision and recall of both TP and TN is equal to 0.99, which means that there are almost no misclassifications of positives/negatives.

Conclusion

Machine learning exhibits huge possibilities of increasing customer satisfaction by solving the problem of delays which cause substantial losses for airlines. The development of algorithms and analytical technologies makes it possible to make a full analysis of the reasons for delays. Modeling allows airlines to be proactive about delay problems to increase efficiency and reduce costs as well as improve customer journey. This strategy not only helps to solve the problem of delays but also increases smoothness and reliability of flight which positively impacts customer satisfaction. Through the analysis of sophisticated interconnections in large heterogeneous data, machine learning models detect key predictors helping airlines to be proactive in order to decrease the number of delays.

REFERENCES

[1] W. Baswardono, D. Kurniadi, A. Mulyani, and D. M. Arifin, "Comparative analysis of decision tree algorithms: Random forest and C4.5 for airlines customer satisfaction classification," *J. Phys.: Conf. Ser.*, vol. 1402, no. 6, Art. no. 066055, 2019, doi: 10.1088/1742-6596/1402/6/066055.

[2] S. Chen, "Estimating customer lifetime value using machine learning techniques," in *Data Mining*. London, U.K.: IntechOpen, 2018, doi: 10.5772/intechopen.76990.

[3] R. Dawes, "Customer expectation, satisfaction and loyalty relationship in Turkish airline industry," *Int. J. Mark. Stud.*, vol. 6, no. 1, pp. 66–74, 2014, doi: 10.5539/ijms.v6n1p66.

[4] D. Gu, S. Khan, I. U. Khan, and S. U. Khan, "Understanding mobile tourism shopping in Pakistan: An integrating framework of innovation diffusion theory and technology acceptance model," *Mobile Inf. Syst.*, vol. 2019, Art. no. 1490617, 2019, doi: 10.1155/2019/1490617.

[5] S. Hong, B. Kim, and Y.-G. Jang, "Correlation analysis of airline customer satisfaction using random forest with deep neural network and support vector machine model," *Int. J. Internet Broadcast. Commun.*, vol. 12, no. 4, pp. 26–32, 2020.

[6] K. Hulliyah, "Predicting airline passenger satisfaction with classification algorithms," *Int. J. Informatics Inf. Syst.*, vol. 4, no. 1, pp. 82–94, 2021, doi: 10.47738/ijiis.v4i1.80.

[7] C. M. Kei, "A study of customer satisfaction in airlines," *Int. J. Trade Econ. Finance*, vol. 10, no. 6, pp. 144–148, 2019, doi: 10.18178/ijtef.2019.10.6.652.

[8] S. Kumar and M. Zymbler, "A machine learning approach to analyze customer satisfaction from airline tweets," *J. Big Data*, vol. 6, no. 1, pp. 1–16, 2019, doi: 10.1186/s40537-019-0224-1.

[9] S. H. Park, M. Y. Kim, Y. J. Kim, and Y. H. Park, "A deep learning approach to analyze airline customer propensities: The case of South Korea," *Appl. Sci.*, vol. 12, no. 4, Art. no. 1916, 2022, doi: 10.3390/app12041916.

[10] J. Prima *et al.*, "Comparison of classification algorithm in classifying airline passenger satisfaction," *J. Sistem Komputer*, vol. 6, no. 1, pp. 92–98, 2022.

[11] H. A. Rady, "Measuring airline service quality using AIRQUAL model: A study applied to EgyptAir," *Int. J. Heritage, Tourism Hosp.*, vol. 12, no. 1, pp. 271–301, 2018, doi: 10.21608/ijhth.2018.31517.

[12] F. A. Shah, Z. Syed, A. Imam, and A. Raza, "The impact of airline service quality on passengers' behavioral intentions using passenger satisfaction as a mediator," *J. Air Transp. Manag.*, vol. 85, Art. no. 101815, 2020, doi: 10.1016/j.jairtraman.2020.101815.

[13] X. Sun, S. Wandelt, C. Zheng, and A. Zhang, "COVID-19 pandemic and air transportation: Successfully navigating the paper hurricane," *J. Air Transp. Manag.*, vol. 94, Art. no. 102062, 2021, doi: 10.1016/j.jairtraman.2021.102062.

[14] Y. Sung and H. Park, "The effect of communication of service employee on customer satisfaction and reuse intention," *J. Econ. Mark. Manag.*, vol. 9, no. 2, pp. 21–31, 2021, doi: 10.20482/jemm.2021.9.2.21.

[15] J. Yimga, "The airline on-time performance impacts of the COVID-19 pandemic," *Transp. Res. Interdiscip. Perspect.*, vol. 10, Art. no. 100386, 2021, doi: 10.1016/j.trip.2021.100386.

[16] C. Zhang and Y. Lu, "Study on artificial intelligence: The state of the art and future prospects," *J. Ind. Inf. Integr.*, vol. 23, Art. no. 100224, 2021, doi: 10.1016/j.jii.2021.100224.

[17] J. Zhou, A. H. Gandomi, F. Chen, and A. Holzinger, "Evaluating the quality of machine learning explanations: A survey on methods and metrics," *Electronics*, vol. 10, no. 5, pp. 1–19, 2021, doi: 10.3390/electronics10050593.